# GENIE 17.0-public release notes

Sage Bionetworks

January 02, 2025

## Release notes

These release notes will encompass the changes made in the 17 consortium series and general center information.

- 14978 samples have been added since the last public release (16.1-public).
- cfDNA samples are now available in the public release.
- Added new validation flags:
  - INT_DOD must be >= INT_CONTACT (github PR)
  - YEAR_DEATH must be >= YEAR_CONTACT (github PR)
- Updated to map race, ethnicity, and sex code mappings from NAACCR v16 to v24 and make granular mappings available for the release files (github PR)
  - Note to expect nulls in the granular mapping columns for SCI and MDA due to invalid clinical files so their files were not processed.
- Fixed a bug creating bed file duplicates for every center (github PR)

## Data Concerns/Issues

These are the known data issues for this release. **Note**: There could be more undiscovered issues.

- The number of expected genes in certain gene panels may be incorrect due to sites submitting invalid gene symbols, coordinates, or incorrect assay information.
- **CRUK**
  - Confirmed genomic data for **CRUK-TS** samples contain no artifacts.
- **CHOP**
  - Fusion panel includes genes that are not on the other panels/pipelines that generate the SNV and indel calls. No BED file for fusion panel.
  - Confirmed genomic data for **CHOP-STNGS** samples contain no artifacts.
- **MSK**
  - Confirmed genomic data for MSK HEME panel samples contain no artifacts.
- **DFCI**
  - Expected discrepancies in expected and submitted gene count because of regions that are targeted for calling structural rearrangements.
- **UHN**
  - Confirmed genomic data for **UHN-54-V1** samples contain no artifacts.
  - UHN has some duplicate bed regions in their processed data. This is a processing issue, no action is needed from sites.
- **YALE**
  - Intentionally reports only amplifications in copy number data due to internal policy.
  - **YALE-OCP-V2** panel's expected gene count is 134 because one intron is included. Gene panels are created with only exons hence the difference in actual vs expected gene counts.
- Foundation Medicine genomic regions discrepancy
  - **DUKE/WAKE** use the Foundation Medicine T5A, T7, DX1, R2D2 bait sets, but some of the bed files uploaded don't seem to match the expected gene count per panel.
- Not all variants have variant counts (t_depth, t_alt_count, t_ref_count).

- Duplicated variants listed from **VICC** are expected due to providing Tempus samples.
- Genome Nexus related issues:
  - Duplicated variants